

ON THE EXPONENTIAL OF SEMI-INFINITE QUASI-TOEPLITZ MATRICES*

DARIO A. BINI[†] AND BEATRICE MEINI[‡]

Abstract. Let $a(z) = \sum_{i \in \mathbb{Z}} a_i z^i$ be a complex valued function defined for $|z| = 1$, such that $\sum_{i \in \mathbb{Z}} |a_i| < \infty$, and let $E = (e_{i,j})_{i,j \in \mathbb{Z}^+}$ be such that $\sum_{i,j \in \mathbb{Z}^+} |e_{i,j}| < \infty$. A semi-infinite quasi-Toeplitz matrix is a matrix of the kind $A = T(a) + E$, where $T(a) = (t_{i,j})_{i,j \in \mathbb{Z}^+}$ is the semi-infinite Toeplitz matrix associated with the symbol $a(z)$, that is, $t_{i,j} = a_{j-i}$ for $i, j \in \mathbb{Z}^+$. We analyze theoretical and computational properties of the exponential of A . More specifically, it is shown that $\exp(A) = T(\exp(a)) + F$ where $F = (f_{i,j})_{i,j \in \mathbb{Z}^+}$ is such that $\sum_{i,j \in \mathbb{Z}^+} |f_{i,j}|$ is finite, i.e., $\exp(A)$ is a semi-infinite quasi-Toeplitz matrix as well, and an effective algorithm for its computation is given. These results can be extended from the function $\exp(z)$ to any function $f(z)$ satisfying mild conditions, and can be applied to finite quasi-Toeplitz matrices.

Key words. Matrix exponential, matrix functions, Taylor series, Toeplitz matrices.

1. The problem and its motivation. Let $a = \{a_k\}_{k \in \mathbb{Z}}$ be a bi-infinite sequence of complex numbers, where the index k ranges in the set \mathbb{Z} of the relative integers, and define $T(a) = (t_{i,j})_{i,j \in \mathbb{Z}^+}$ the semi-infinite Toeplitz matrix such that $t_{i,j} = a_{j-i}$ for i, j in the set \mathbb{Z}^+ of positive integers.

In this paper, we analyze the problem of computing the matrix exponential of semi-infinite matrices of the form

$$A = T(a) + E$$

where $E = (e_{i,j})_{i,j \in \mathbb{Z}^+}$ is such that $\sum_{i,j \in \mathbb{Z}^+} |e_{i,j}| < +\infty$ and $\sum_{k \in \mathbb{Z}} |ka_k| < +\infty$. We refer to this class of matrices as Quasi-Toeplitz, in short, QT-matrices.

The problem of computing the matrix exponential of Toeplitz and quasi-Toeplitz matrices is encountered in diverse applications, like the Erlangian approximation of Markovian fluid queues [10], [5], or the discretization of integro-differential equations with a shift-invariant kernel which describe the pricing of single-asset options modeled by jump-diffusion processes [16], [17], [19], where matrices are finite but have huge dimensions since their size has an asymptotic meaning. Another simple example comes from the numerical solution of the heat equation $\frac{\partial u}{\partial t} = \gamma \frac{\partial^2 u}{\partial x^2}$ where the second derivative in space is discretized with the three-point finite difference formula, so that the equation is reduced to an ordinary differential equation of the kind $v' = Av$, being v the vector function of the values of $u(x, t)$ at the discretization nodes. If the spatial domain is infinite, then $A = -\frac{\gamma}{h^2} \text{trid}(-1, 2, -1)$ is a semi-infinite tridiagonal Toeplitz matrix and the solution can be expressed as $v(t) = \exp(tA)v(0)$.

In other problems, like the analysis of random walks in the quarter plane [11], [26], or in the tandem Jackson queue [23], [20], where the number of states is infinite denumerable, one encounters semi-infinite probability matrices which are block Toeplitz where the blocks have the form $T(a) + E$, with $T(a)$ banded Toeplitz and E having a finite number of nonzero entries.

These applications make it interesting to design effective algorithms for computing the exponential, and other analytic functions, of matrices which are quasi-Toeplitz.

*The research was carried out with the support of GNCS of INdAM

[†]Dipartimento di Matematica, Università di Pisa, Largo Bruno Pontecorvo 5, 56127 Pisa (dario.bini@unipi.it)

[‡]Dipartimento di Matematica, Università di Pisa, Largo Bruno Pontecorvo 5, 56127 Pisa (beatrice.meini@unipi.it)

Due to its importance, the problem of computing the exponential of a Toeplitz matrix has been analyzed in several papers among which [16], [18], [21], [27] where different techniques, like using Krylov subspaces or displacement operators, are used effectively. The analysis of the exponential of infinite matrices is considered in [12], [15], while truncating to finite size the exponential of a general infinite matrix is considered in [25]. Solving certain matrix polynomial equations with QT matrices as coefficients has been recently considered in [7].

1.1. New results. In this paper, we provide some theoretical results and some algorithmic advances concerning the computation of the exponential of infinite Quasi-Toeplitz matrices $A = T(a) + E$ where $a(z) = \sum_{i \in \mathbb{Z}} a_i z^i$ is such that $\sum_{i \in \mathbb{Z}} |a_i|$ is finite. That is, $a(z)$ and $a'(z) = \sum_{i \in \mathbb{Z}} i a_i z^{i-1}$ belong to the Wiener class

$$\mathcal{W} = \left\{ a(z) = \sum_{i \in \mathbb{Z}} a_i z^i : \sum_{i \in \mathbb{Z}} |a_i| < \infty \right\}.$$

The approach that we present is general and can be easily applied to the finite case and extended to the computation of more general analytic functions. We rely on the approximation of $\exp(A)$ given by the Taylor series truncation $S_k = \sum_{i=0}^k \frac{1}{i!} A^i$, for a sufficiently large k . We exploit the Toeplitz structure of A and of its powers by representing S_k as $S_k = T(s_k) + F_k$, where $s_k(z) = \sum_{i=0}^k \frac{1}{i!} a(z)^i$ is the Taylor series truncation of $\exp(a)$ and F_k is a suitable correction.

More precisely, we prove that if the function $a(z)$ is such that $a(z), a'(z) \in \mathcal{W}$, then the matrix exponential $S = \exp(A)$ is well defined by its power series expansion $\exp(A) = \sum_{i=0}^{\infty} \frac{1}{i!} A^i$ and is still of the form $T(b) + F$ where $b(z) = \sum_{i \in \mathbb{Z}} b_i z^i = \exp(a(z))$ and $\sum_{i,j \in \mathbb{Z}^+} |f_{i,j}| < +\infty$. That is, $\exp(A)$ has a *Toeplitz component* and a *correction component*. The former, represented by the coefficients b_k , $k \in \mathbb{Z}$, can be easily approximated to any precision by applying the evaluation and interpolation technique at the roots of the unity to the equation $b(z) = \exp(a(z))$. The correction F can be approximated to any precision by means of an easily computable recurrence which generates a sequence of matrices F_k such that $\lim_{k \rightarrow \infty} F_k = F$.

This representation of a function of a Toeplitz matrix differs from those given in [27], [16], [18] [21] which rely on the displacement operator and displacement rank. The idea of representing the infinite matrix $\exp(A)$ as a Toeplitz part associated with a suitable function plus a correction having a finite sum of the moduli of its entries, is fundamental to deal with *infinite* matrices by using only a *finite* number of parameters. In fact, the coefficients g_i of a function $g(z) = \sum_{i \in \mathbb{Z}} g_i z^i$ such that $\sum_{i \in \mathbb{Z}} |i g_i| < \infty$, decay to zero as $i \rightarrow \pm\infty$ so that $g(z)$ can be approximated by means of a Laurent polynomial $\hat{g}(z) = \sum_{i=-n_-}^{n_+} g_i z^i$, for $n_-, n_+ > 0$ sufficiently large. Moreover, the finiteness of the sum $\sum_{i,j \in \mathbb{Z}^+} |f_{i,j}|$ allows one to approximate F with a finite matrix generally of small rank.

The decomposition $A = T(a) + E$ is quite natural and widely used in the analysis of spectral properties of sequences $\{A_n\}$ of $n \times n$ Toeplitz-like matrices where the additive decomposition is given as the sum of a Toeplitz part, plus a matrix of small rank plus a matrix of small norm. We refer the reader to [4], [24] for the basic properties and for a list of references in this regard. The same kind of decomposition is described in [9, Example 2.28] where, unlike in this paper, it is assumed that the correction E is a compact operator in ℓ^2 . The boundedness of the operator norm $\|E\|_2$ is not enough for our computational goals since it does not imply the boundedness of $\sum_{i,j \in \mathbb{Z}^+} |e_{i,j}|$.

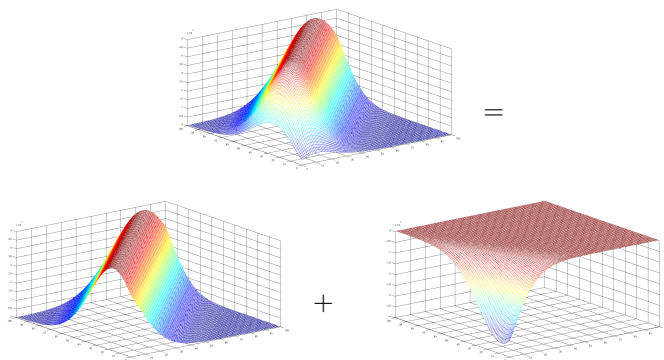


FIGURE 1.1. *Decomposition of the matrix exponential as the sum of a Toeplitz matrix and of a correction with low numerical rank. The matrix is $\exp(T(a))$ where $a(z) = \sum_{i=-5}^{10} z^i$*

The representation of a matrix as sum of a Toeplitz part and a correction is described pictorially in Figure 1.1 where the exponential of $T(a)$ is shown for $a(z) = \sum_{i=-5}^{10} z^i$. The two components $T(\exp(a))$, i.e., the Toeplitz part, and the “compact” correction F are displayed together with their sum.

The same approach can be applied to the finite case, where the correction F involves both the north-west and the south-east corners. In the finite case, the advantage of this representation is much appreciated if the size of the matrix is larger than the rank of the correction.

Our approach is more effective when the exponential of the QT-matrix has a large decay of the band. In fact, in this case, both the degree of the Laurent polynomial which approximates $\exp(a)$ and the rank of the correction matrix F are small. The analysis of the decay of the band of matrix functions, and more specifically of Toeplitz matrices, has been recently performed in the papers [2], [3], [22]. The decay of singular values of matrices with low displacement rank, including positive definite Hankel matrices, has been performed in [1]. An immediate consequence of our results is that the exponential of a quasi-Toeplitz matrix can be approximated up to any error ϵ by a banded matrix.

The acceleration techniques, based on scaling the variable A and squaring the exponential of the scaled matrix, can be applied to reduce the number of terms in the power series expansion needed to reach a sufficiently accurate approximation. Similarly, our approach can be combined with Padé approximation to reduce the cost of the computation.

The algorithm that we have obtained this way has been implemented in Matlab and tested with some “synthetic” problems and with some matrices taken from the applications.

From the numerical experiments it turns out that the correction component F such that $\exp(T(a)) = T(\exp(a)) + F$, has generally very low rank. The same holds for the correction E_k such that $T(a)^k = T(a^k) + E_k$. This property, which will be object of our future research analysis, is at the basis of the effectiveness of our algorithm.

Extensions of this approach with the implementation of quasi-Toeplitz matrix arithmetic are given in [7], applications to general analytic functions either expressed as power series or as Cauchy integrals and the algorithmic analysis of the finite case are treated in [6].

The paper is organized as follows. In Section 2 we recall some preliminary results concerning semi-infinite Toeplitz matrices, introduce the norms $\|a\|_{\mathcal{W}} = \sum_{i \in \mathbb{Z}} |a_i|$ and $\|F\|_{\mathcal{F}} = \sum_{i,j \in \mathbb{Z}^+} |f_{i,j}|$, and recall that the linear space \mathcal{F} formed by semi-infinite matrices F such that $\|F\|_{\mathcal{F}}$ is finite is a Banach algebra. In Section 3 we consider the case of a Toeplitz matrix $T(a)$. We prove that if $a(z), a'(z) \in \mathcal{W}$ then $T(a)^i = T(a^i) + E_i$ where $E_i \in \mathcal{F}$ and give an explicit relation between E_i and E_{i-1} . This relation enables us to provide a bound to $\|E_i\|_{\mathcal{F}}$ given in terms of $\|a\|_{\mathcal{W}}$ and of $\|a'\|_{\mathcal{W}}$. In Section 4 we extend these results to the case of a QT matrix $A = T(a) + E$. In Section 5 we describe in detail the algorithm for computing the exponential of a Toeplitz matrix and we outline the case of a QT matrix. Section 6 reports the results of some numerical experiments while Section 7 draws some final remarks and conclusions.

2. Preliminaries. We recall the basic definition of matrix exponential of an $n \times n$ matrix and the main properties of semi-infinite Toeplitz matrices which will be used in our analysis.

For an $n \times n$ matrix A , it is well known that the series $\exp(A) := \sum_{i=0}^{+\infty} \frac{1}{i!} A^i$ is convergent and defines the matrix exponential of A . We refer to the book by N. Higham [14] for the concept of matrix function and for more details on the matrix exponential. Indeed, defining S_k the partial sum

$$S_k = \sum_{i=0}^k \frac{1}{i!} A^i, \quad (2.1)$$

and the remainder R_k of the series as $R_k = \sum_{i=k+1}^{\infty} \frac{1}{i!} A^i$, for any matrix norm $\|\cdot\|$ such that $\|A^2\| \leq \|A\|^2$ it follows that

$$\|R_k\| = \left\| \sum_{i=k+1}^{\infty} \frac{1}{i!} A^i \right\| \leq \sum_{i=k+1}^{\infty} \frac{1}{i!} \|A\|^i \quad (2.2)$$

so that $\lim_{k \rightarrow \infty} \|R_k\| = 0$ which implies the convergence of the sequence S_k .

This property is still valid if $A = (a_{i,j})_{i,j \in \mathbb{Z}^+}$ is a semi-infinite matrix provided that A belongs to a Banach algebra \mathcal{A} , that is an algebra endowed with a sub-multiplicative norm $\|\cdot\|$, such that $\|AB\| \leq \|A\| \cdot \|B\|$ for any $A, B \in \mathcal{A}$, which makes it a Banach space. Indeed, for $A \in \mathcal{A}$, consider the sequence $\{S_k\}_k$ defined in (2.1). For $i > j$ we have

$$\|S_i - S_j\| \leq \sum_{h=j+1}^i \frac{1}{h!} \|A^h\| \leq \sum_{h=j+1}^i \frac{1}{h!} \|A\|^h.$$

From this bound it follows that for any $\epsilon > 0$ there exists $k > 0$ such that $\|S_i - S_j\| \leq \epsilon$ for any $i > j \geq k$. That is, $\{S_k\}_k$ is a Cauchy sequence. Since by definition of Banach space, the Cauchy sequences in \mathcal{A} have a limit in \mathcal{A} , there exists a matrix L such that $\lim_{k \rightarrow \infty} \|S_k - L\| = 0$. We denote $L = \exp(A)$. Thus, the matrix exponential is well defined also for a semi-infinite matrix A provided that it belongs to a Banach algebra with respect to some norm.

We now recall some results concerning infinite Toeplitz matrices. For more details on this topic we refer the reader to the book by Böttcher and Grudsky [8].

Let $\mathcal{W} = \{a(z) = \sum_{i \in \mathbb{Z}} a_i z^i : \sum_{i \in \mathbb{Z}} |a_i| < +\infty\}$ denote the Wiener algebra formed by Laurent power series, defined on the unit circle $\mathbb{T} = \{z \in \mathbb{C} : |z| = 1\}$.

$1\}$, whose coefficients have a finite sum of their moduli. It is well known that \mathcal{W} endowed with the norm $\|a\|_{\mathcal{W}} = \sum_{i \in \mathbb{Z}} |a_i|$ is a Banach algebra. For $a(z) \in \mathcal{W}$ denote $T(a)$ the semi-infinite Toeplitz matrix whose entries $t_{i,j}$ are such that $t_{i,j} = a_{j-i}$ for $i, j \in \mathbb{Z}^+$, where \mathbb{Z}^+ denotes the set of positive integers. Denote also $a_+(z)$ and $a_-(z)$ the power series defined by $a_+(z) = \sum_{i \in \mathbb{Z}^+} a_i z^i$ and $a_-(z) = \sum_{i \in \mathbb{Z}^+} a_{-i} z^i$ so that $a(z) = a_0 + a_+(z) + a_-(z^{-1})$. Finally, given the power series $b(z) = \sum_{z \in \mathbb{Z}^+} b_i z^i$ define $H(b) = (h_{i,j})$ the Hankel matrix such that $h_{i,j} = b_{i+j-1}$, for $i, j \in \mathbb{Z}^+$.

Any semi-infinite matrix $S = (s_{i,j})_{i,j \in \mathbb{Z}^+}$ can be viewed as a linear operator, acting on semi-infinite vectors $v = (v_i)_{i \in \mathbb{Z}^+}$, which maps the vector v onto the vector u such that $u_i = \sum_{j \in \mathbb{Z}^+} s_{i,j} v_j$, provided that the summations are finite. For any $p \geq 1$, included $p = \infty$, we may define the Banach space ℓ^p formed by all the semi-infinite vectors $v = (v_i)_{i \in \mathbb{Z}^+}$ such that $\|v\|_p = (\sum_{i \in \mathbb{Z}^+} |v_i|^p)^{\frac{1}{p}} < \infty$, where for $p = \infty$ we have $\|v\|_{\infty} = \sup_{i \in \mathbb{Z}^+} |v_i|$. It is well known that these norms induce the corresponding operator norms $\|S\|_p = \sup_{\|v\|_p=1} \|Sv\|_p$ which are sub-multiplicative, i.e., $\|AB\|_p \leq \|A\|_p \|B\|_p$ for any semi-infinite matrices A, B with finite ℓ^p norm, and that the linear space formed by the latter semi-infinite matrices forms a Banach algebra.

We may wonder if the matrices $T(a)$, $H(a_+)$ and $H(a_-)$ define linear operators acting on the Banach space ℓ^p and if they have a finite operator norm. The answer to this question is given by the following result of [8] which relates the matrix $T(a)T(b)$ with $T(ab)$, $H(a_-)$ and $H(a_+)$.

THEOREM 2.1. *For $a(z), b(z) \in \mathcal{W}$ let $c(z) = a(z)b(z)$. Then we have*

$$T(a)T(b) = T(c) - H(a_-)H(b_+).$$

Moreover, for any $p \geq 1$, including $p = \infty$, we have

$$\|T(a)\|_p \leq \|a\|_{\mathcal{W}}, \quad \|H(a_-)\|_p \leq \|a_-\|_{\mathcal{W}}, \quad \|H(b_+)\|_p \leq \|b_+\|_{\mathcal{W}}.$$

The above result implies that the product of two Toeplitz matrices can be written as a Toeplitz matrix plus a correction whose ℓ^p -norm is bounded by $\|a_-\|_{\mathcal{W}} \|b_+\|_{\mathcal{W}} \leq \|a\|_{\mathcal{W}} \|b\|_{\mathcal{W}}$.

For a semi-infinite matrix S we introduce the norm $\|S\|_{\mathcal{F}} = \sum_{i,j \in \mathbb{Z}^+} |s_{i,j}|$, which coincides with the ℓ^1 norm if we consider the matrix S as the vector $s_{i,j}$ where the pairs (i, j) are ordered in a triangular fashion starting from the leftmost top corner $(1, 1)$. We denote by \mathcal{F} the set of semi-infinite matrices such that $\|S\|_{\mathcal{F}} < \infty$. The next result states that $\|\cdot\|_{\mathcal{F}}$ is sub-multiplicative in \mathcal{F} and that \mathcal{F} is a Banach algebra.

THEOREM 2.2. *For $A, B \in \mathcal{F}$ it holds $\|AB\|_{\mathcal{F}} \leq \|A\|_{\mathcal{F}} \cdot \|B\|_{\mathcal{F}}$, moreover, \mathcal{F} is a Banach algebra.*

Proof. Let $C = AB$ so that $c_{i,j} = \sum_{r \in \mathbb{Z}^+} a_{i,r} b_{r,j}$. Since $\sum_{i,r \in \mathbb{Z}^+} |a_{i,r}| < +\infty$, then $\alpha_i = \sum_{r \in \mathbb{Z}^+} |a_{i,r}| < +\infty$ and $\sum_{i \in \mathbb{Z}^+} \alpha_i = \|A\|_{\mathcal{F}}$. Similarly, $\beta_j = \sum_{r \in \mathbb{Z}^+} |b_{r,j}| < +\infty$ and $\sum_{j \in \mathbb{Z}^+} \beta_j = \|B\|_{\mathcal{F}}$. Whence we obtain

$$|c_{i,j}| \leq \sum_{r \in \mathbb{Z}^+} |a_{i,r} b_{r,j}| \leq \alpha_i \beta_j$$

since, in general $\sum_{r \in \mathbb{Z}^+} x_r y_r \leq (\sum_{r \in \mathbb{Z}^+} x_r)(\sum_{r \in \mathbb{Z}^+} y_r)$ for any $x_r, y_r \geq 0$ such that $\sum_{r \in \mathbb{Z}^+} x_r$ and $\sum_{r \in \mathbb{Z}^+} y_r$ are bounded. Thus we get

$$\sum_{i,j \in \mathbb{Z}^+} |c_{i,j}| \leq \sum_{i \in \mathbb{Z}^+} \alpha_i \sum_{j \in \mathbb{Z}^+} \beta_j = \|A\|_{\mathcal{F}} \cdot \|B\|_{\mathcal{F}}.$$

In order to prove completeness of \mathcal{F} , observe that the norm $\|\cdot\|_{\mathcal{F}}$ corresponds to the ℓ^1 norm in the space of infinite sequences having finite sum of their moduli. This way, the space \mathcal{F} actually coincides with ℓ^1 , which is a Banach space. \square

The following result will be useful to prove boundedness properties of the exponential of Toeplitz matrices.

LEMMA 2.3. *If $E = (e_{i,j}) \in \mathcal{F}$ and $a(z) \in \mathcal{W}$, then $\|T(a)E\|_{\mathcal{F}} \leq \|a\|_{\mathcal{W}}\|E\|_{\mathcal{F}}$.*

Proof. Let $V = T(a)E$ so that $v_{i,j} = \sum_{r \in \mathbb{Z}^+} a_{r-i} e_{r,j}$. Observe that for any $j, r \in \mathbb{Z}^+$, one has $\sum_{i \in \mathbb{Z}^+} |a_{r-i} e_{r,j}| \leq \sum_{k \in \mathbb{Z}} |a_k| \cdot |e_{r,j}| = \|a\|_{\mathcal{W}} |e_{r,j}|$. From this inequality we find that

$$\begin{aligned} \|V\|_{\mathcal{F}} &= \sum_{i,j \in \mathbb{Z}^+} |v_{i,j}| \leq \sum_{i,j \in \mathbb{Z}^+} \sum_{r \in \mathbb{Z}^+} |a_{r-i} e_{r,j}| = \sum_{r,j \in \mathbb{Z}^+} \sum_{i \in \mathbb{Z}^+} |a_{r-i} e_{r,j}| \\ &\leq \|a\|_{\mathcal{W}} \sum_{r,j \in \mathbb{Z}^+} |e_{r,j}| = \|a\|_{\mathcal{W}} \|E\|_{\mathcal{F}}. \end{aligned}$$

\square

3. Exponential of a semi-infinite Toeplitz matrix. In this section we study the properties of the exponential of a semi-infinite Toeplitz matrix, by relating in particular $\exp(T(a))$ to $T(\exp(a))$.

Let $a \in \mathcal{W}$ and consider the associated semi-infinite Toeplitz matrix $T(a)$. Since, for Theorem 2.1, $T(a)$ belong to the Banach algebra of linear operators over ℓ^p , then $\exp(T(a))$ is well defined. Moreover, we may write

$$\exp(T(a)) = \sum_{i=0}^{\infty} \frac{1}{i!} T(a)^i.$$

From Theorem 2.1 and for the monotonicity of the function $\exp(z)$ we have

$$\|\exp(T(a))\|_p \leq \sum_{i=0}^{\infty} \frac{1}{i!} \|T(a)\|_p^i = \exp(\|T(a)\|_p) \leq \exp(\|a\|_{\mathcal{W}}).$$

Now we will take a closer look at $\exp(T(a))$ and relate it to $T(\exp(a))$. Since \mathcal{W} is a Banach algebra, the function $\exp(z)$ is well defined over \mathcal{W} and we have

$$\exp(a(z)) = \sum_{i=0}^{+\infty} \frac{1}{i!} a(z)^i.$$

We first relate $T(a)^i$ to $T(a^i)$, for $i \geq 2$. From Theorem 2.1 we may write $T(a)^2 = T(a^2) + E_2$, where $E_2 = -H(a_-)H(a_+)$. For a general $i \geq 0$ define E_i as

$$E_i = T(a)^i - T(a^i), \quad (3.1)$$

where $E_0 = 0$, $E_1 = 0$. Then we have the following

THEOREM 3.1. *Let $a \in \mathcal{W}$ and let $E_i = T(a)^i - T(a^i)$, for $i \geq 1$. Then*

$$\begin{aligned} E_i &= T(a)E_{i-1} - H(a_-)H((a^{i-1})_+), \quad i \geq 2, \\ E_1 &= 0. \end{aligned} \quad (3.2)$$

Moreover, for any $i \geq 1$ and any integer $p \geq 1$, included $p = \infty$,

$$\|E_i\|_p \leq (i-1)\|a\|_{\mathcal{W}}^i. \quad (3.3)$$

Proof. From the equation $T(a)^i = T(a)T(a)^{i-1}$ and from Theorem 2.1 we obtain

$$\begin{aligned} T(a)^i &= T(a)T(a)^{i-1} = T(a)[T(a^{i-1}) + E_{i-1}] \\ &= T(a^i) - H(a_-)H((a^{i-1})_+) + T(a)E_{i-1} \\ &= T(a^i) + E_i, \end{aligned}$$

with $E_i = -H(a_-)H((a^{i-1})_+) + T(a)E_{i-1}$. Whence we deduce recurrence (3.2). Moreover, for any ℓ^p -norm, since $\|a_+\|_{\mathcal{W}} \leq \|a\|_{\mathcal{W}}$, $\|a_-\|_{\mathcal{W}} \leq \|a\|_{\mathcal{W}}$ and $\|a^i\|_{\mathcal{W}} \leq \|a\|_{\mathcal{W}}^i$, applying once again Theorem 2.1, from (3.2) we obtain

$$\|E_i\|_p \leq \|a\|_{\mathcal{W}}\|E_{i-1}\|_p + \|a\|_{\mathcal{W}}\|a^{i-1}\|_{\mathcal{W}} \leq \|a\|_{\mathcal{W}}\|E_{i-1}\|_p + \|a\|_{\mathcal{W}}^i.$$

By using an induction argument we arrive at the bound (3.3). \square

Now define S_k , F_k and G_k as follows

$$\begin{aligned} S_k &= \sum_{i=0}^k \frac{1}{i!} T(a)^i = G_k + F_k \\ G_k &= \sum_{i=0}^k \frac{1}{i!} T(a^i), \quad F_k = \sum_{i=0}^k \frac{1}{i!} E_i. \end{aligned}$$

Observe that $G_k = \sum_{i=0}^k \frac{1}{i!} T(a^i) = T(\sum_{i=0}^k \frac{1}{i!} a^i)$ is such that $\lim_{k \rightarrow \infty} G_k = T(\exp(a))$. Thus, since $\lim_{k \rightarrow \infty} S_k = \exp(T(a))$, then there exists the limit

$$\lim_{k \rightarrow \infty} F_k = \exp(T(a)) - T(\exp(a)) =: F. \quad (3.4)$$

This way, we may write the exponential of a semi-infinite Toeplitz matrix $T(a)$ associated with a symbol $a \in \mathcal{W}$ in the form

$$\exp(T(a)) = T(\exp(a)) + F$$

where $F = \sum_{i=0}^{\infty} \frac{1}{i!} E_i$ has ℓ^p norm bounded by

$$\begin{aligned} \|F\|_p &\leq \|\exp(T(a))\|_p + \|T(\exp(a))\|_p \leq \exp(\|a\|_{\mathcal{W}}) + \|\exp(a)\|_{\mathcal{W}} \\ &\leq 2 \exp(\|a\|_{\mathcal{W}}) \end{aligned} \quad (3.5)$$

since $\|\exp(a)\|_{\mathcal{W}} \leq \exp(\|a\|_{\mathcal{W}})$.

Now, our next step is to prove a stronger property. That is, we show that under certain conditions, the correction F is such that $\|F\|_{\mathcal{F}} < \infty$. This property is stronger than $\|F\|_p < \infty$ and is very useful from the computational point of view since it implies that for any $\epsilon > 0$ there exists k such that $|f_{i,j}| \leq \epsilon$ for any $i, j > k$. This bound allows us to represent F , up to within any given error bound, by using a finite number of parameters.

Let $a(z) \in \mathcal{W}$ be such that $a'(z) \in \mathcal{W}$ where $a'(z) = \sum_{i \in \mathbb{Z}} i a_i z^{i-1}$, so that $\sum_{i \in \mathbb{Z}} |a_i|$ and $\sum_{i \in \mathbb{Z}} |i a_i|$ are finite. Moreover, observe that if $a(z), a'(z) \in \mathcal{W}$ and $b(z), b'(z) \in \mathcal{W}$, then also $c(z), c'(z) \in \mathcal{W}$ for $c(z) = a(z) + b(z)$ and for $c(z) = a(z)b(z)$. This property enables us to prove the following

THEOREM 3.2. *Let $a(z), a'(z) \in \mathcal{W}$. Then for any $k \geq 1$ we have*

$$\|H(a_-)H((a^{k-1})_+)\|_{\mathcal{F}} \leq (k-1)\|a\|_{\mathcal{W}}^{k-2}\|a'\|_{\mathcal{W}}^2.$$

Proof. We have $\|H(a_-)\|_{\mathcal{F}} = \sum_{i,j \in \mathbb{Z}^+} |a_{1-i-j}| = \sum_{h \in \mathbb{Z}^+} h|a_{-h}| \leq \|a'\|_{\mathcal{W}}$ which is finite since $a'(z) \in \mathcal{W}$. Similarly, $\|H((a^{k-1})_+)\|_{\mathcal{F}} \leq \|(a^{k-1})'\|_{\mathcal{W}} < \infty$ since both the functions $a^{k-1}(z)$ and $(a^{k-1}(z))'$ belong to \mathcal{W} . Thus, for the matrix product $L_k = H(a_-)H((a^{k-1})_+)$ we find that

$$\|L_k\|_{\mathcal{F}} \leq \|H(a_-)\|_{\mathcal{F}} \cdot \|H((a^{k-1})_+)\|_{\mathcal{F}} \leq \|a'\|_{\mathcal{W}} \|(a^{k-1})'\|_{\mathcal{W}}.$$

Now, since $(a^{k-1}(z))' = (k-1)a^{k-2}(z)a'(z)$, we have

$$\|(a^{k-1})'\|_{\mathcal{W}} \leq (k-1)\|a\|_{\mathcal{W}}^{k-2}\|a'\|_{\mathcal{W}}.$$

Thus we get $\|L_k\|_{\mathcal{F}} \leq (k-1)\|a\|_{\mathcal{W}}^{k-2}\|a'\|_{\mathcal{W}}^2$. \square

From the above result and from Lemma 2.3, we deduce the main theorem of this section

THEOREM 3.3. *Under the assumptions of Theorem 3.2, for the matrices E_i of (3.1) and $F = \lim_k F_k$, with $F_k = \sum_{i=0}^k \frac{1}{i!} E_i$, we have*

$$\begin{aligned} \|E_i\|_{\mathcal{F}} &\leq \frac{i(i-1)}{2} \|a'\|_{\mathcal{W}}^2 \|a\|_{\mathcal{W}}^{i-2}, \quad i \geq 0, \\ \|F\|_{\mathcal{F}} &\leq \frac{1}{2} \|a'\|_{\mathcal{W}}^2 \exp(\|a\|_{\mathcal{W}}). \end{aligned}$$

Proof. From Theorem 3.2 and from (3.2) we have

$$\|E_i\|_{\mathcal{F}} \leq \|T(a)E_{i-1}\|_{\mathcal{F}} + (i-1)\|a\|_{\mathcal{W}}^{i-2}\|a'\|_{\mathcal{W}}^2, \quad i \geq 2,$$

where $E_0 = E_1 = 0$. From Lemma 2.3 we deduce that

$$\|E_i\|_{\mathcal{F}} \leq \|a\|_{\mathcal{W}} \|E_{i-1}\|_{\mathcal{F}} + (i-1)\|a\|_{\mathcal{W}}^{i-2}\|a'\|_{\mathcal{W}}^2.$$

Therefore, by using the induction argument we arrive at

$$\|E_i\|_{\mathcal{F}} \leq \frac{i(i-1)}{2} \|a'\|_{\mathcal{W}}^2 \|a\|_{\mathcal{W}}^{i-2}, \quad i \geq 2,$$

which proves the first bound. This implies that

$$\|F\|_{\mathcal{F}} \leq \sum_{i=0}^{\infty} \frac{1}{i!} \|E_i\|_{\mathcal{F}} \leq \frac{1}{2} \|a'\|_{\mathcal{W}}^2 \exp(\|a\|_{\mathcal{W}}),$$

which completes the proof. \square

This way, we find that $\exp(T(a)) = T(\exp(a)) + F$, where F is such that $\|F\|_{\mathcal{F}} < +\infty$. In different words, the exponential of a semi-infinite Toeplitz matrix is a semi-infinite Toeplitz matrix, up to a correction belonging to the set \mathcal{F} . This property, combined with equation (3.2), enables us to design algorithms for computing the exponential of a Toeplitz matrix expressed in the form $T(\exp(a)) + F$. This is possible since both $\exp(a)$ and F can be represented up to an arbitrarily small error, by means of a finite number of parameters. We will see the design and analysis of this class of algorithms in the next Section 5.

Observe also that while the relation $F = \exp(T(a)) - T(\exp(a))$ is useful to provide the upper bound $\|F\|_p \leq \|\exp(T(a))\|_p + \|T(\exp(a))\|_p \leq 2 \exp(\|a\|_{\mathcal{W}})$ as we did in

(3.5), it cannot be used to provide finite upper bounds to $\|F\|_{\mathcal{W}}$ since any nonzero Toeplitz matrix $T(a)$ has infinite \mathcal{W} norm.

Before dealing with the algorithmic aspects of this problem, we wish to extend our analysis to the case of a matrix $A = T(a) + E$ where $\|E\|_{\mathcal{F}} < +\infty$. Also in this case we will prove that $\exp(A)$ can be written in the form $T(\exp(a)) + D$ where $\|D\|_{\mathcal{F}} < +\infty$. Matrices of the kind $T(a) + E$ are encountered in applications related to the analysis of certain stochastic processes.

4. A generalization. Let $A = T(a) + E$, where $a(z), a'(z) \in \mathcal{W}$, and E is any semi-infinite matrix belonging to the set \mathcal{F} . The argument used in Section 3 can be applied to provide an expression to $\exp(A)$. In fact, we may write $\exp(A) = S_k + R_k$ where $S_k = \sum_{i=0}^k \frac{1}{i!} A^i$ and $R_k = \sum_{i=k+1}^{\infty} \frac{1}{i!} A^i$, so that $S_k = G_k + F_k$ with

$$G_k = \sum_{i=0}^k \frac{1}{i!} T(a^i), \quad F_k = \sum_{i=0}^k \frac{1}{i!} D_i,$$

where

$$D_i = A^i - T(a^i), \quad i \geq 0. \quad (4.1)$$

Observe that $D_0 = 0$, $D_1 = E$. As in the previous section, there exists

$$F = \lim_{k \rightarrow \infty} F_k = \lim_{k \rightarrow \infty} S_k - \lim_{k \rightarrow \infty} G_k = \exp(A) - T(\exp(a)).$$

Our goal is to estimate $\|F\|_{\mathcal{F}}$ and to show that $\|F\|_{\mathcal{F}} < \infty$.

From (4.1) we deduce that, for $i \geq 1$,

$$\begin{aligned} A^i &= (T(a) + E)(T(a^{i-1}) + D_{i-1}) \\ &= T(a)T(a^{i-1}) + ET(a^{i-1}) + AD_{i-1} \end{aligned}$$

and, in view of Theorem 2.1, it follows that

$$A^i = T(a^i) - H(a_-)H((a^{i-1})_+) + ET(a^{i-1}) + AD_{i-1}.$$

Hence, we obtain

$$D_i = AD_{i-1} - H(a_-)H((a^{i-1})_+) + ET(a^{i-1}). \quad (4.2)$$

In view of Lemma 2.3, we have $\|T(a)D_{i-1}\|_{\mathcal{F}} \leq \|a\|_{\mathcal{W}}\|D_{i-1}\|_{\mathcal{F}}$ so that

$$\|AD_{i-1}\|_{\mathcal{F}} = \|T(a)D_{i-1} + ED_{i-1}\|_{\mathcal{F}} \leq (\|a\|_{\mathcal{W}} + \|E\|_{\mathcal{F}})\|D_{i-1}\|_{\mathcal{F}}.$$

Therefore, from the above inequality and from Theorem 3.2, in view of (4.2), we obtain

$$\begin{aligned} \|D_i\|_{\mathcal{F}} &\leq (\|a\|_{\mathcal{W}} + \|E\|_{\mathcal{F}})\|D_{i-1}\|_{\mathcal{F}} + (i-1)\|a\|_{\mathcal{W}}^{i-2}\|a'\|_{\mathcal{W}}^2 + \|a\|_{\mathcal{W}}^{i-1}\|E\|_{\mathcal{F}}, \\ &\leq \xi\|D_{i-1}\|_{\mathcal{F}} + \gamma_i, \quad i \geq 1, \end{aligned}$$

where

$$\xi = \|a\|_{\mathcal{W}} + \|E\|_{\mathcal{F}}, \quad \gamma_i = (i-1)\|a\|_{\mathcal{W}}^{i-2}\|a'\|_{\mathcal{W}}^2 + \|a\|_{\mathcal{W}}^{i-1}\|E\|_{\mathcal{F}}, \quad i \geq 1, \quad (4.3)$$

and $\|D_0\|_{\mathcal{F}} = 0$, $\|D_1\|_{\mathcal{F}} = \|E\|_{\mathcal{F}}$. Thus we may bound $\|D_i\|_{\mathcal{F}}$ with the value that the polynomial $p(z) = \sum_{j=0}^{i-1} z^j \gamma_{i-j}$ takes at $\xi = \|a\|_{\mathcal{W}} + \|E\|_{\mathcal{F}}$, i.e.,

$$\|D_i\|_{\mathcal{F}} \leq \sum_{j=0}^{i-1} \xi^j \gamma_{i-j}, \quad \xi = \|a\|_{\mathcal{W}} + \|E\|_{\mathcal{F}}. \quad (4.4)$$

Thus, concerning the sequence F_k , from (4.4) we obtain

$$\|F_k\|_{\mathcal{F}} \leq \sum_{i=1}^k \frac{1}{i!} \|D_i\|_{\mathcal{F}} \leq \sum_{i=1}^k \sum_{j=0}^{i-1} \frac{1}{i!} \xi^j \gamma_{i-j}.$$

For notational simplicity set $\alpha = \|a\|_{\mathcal{W}}$, $\beta = \|E\|_{\mathcal{F}}$ so that $\xi = \alpha + \beta$ and $\gamma_k = (k-1)\alpha^{k-2}\|a'\|_{\mathcal{W}}^2 + \alpha^{k-1}\beta$. Then, since $\alpha \leq \xi$, we have $\gamma_k \leq (k-1)\xi^{k-2}\|a'\|_{\mathcal{W}}^2 + \xi^{k-1}\beta$. Whence we deduce that

$$\begin{aligned} \|F_k\|_{\mathcal{F}} &\leq \sum_{i=1}^k \sum_{j=0}^{i-1} \frac{1}{i!} [\|a'\|_{\mathcal{W}}^2 (i-j-1) \xi^{i-2} + \beta \xi^{i-1}] \\ &= \|a'\|_{\mathcal{W}}^2 \sum_{i=1}^k \frac{1}{2} \frac{i(i-1)}{i!} \xi^{i-2} + \beta \sum_{i=1}^k \frac{i}{i!} \xi^{i-1} \\ &\leq \frac{1}{2} \|a'\|_{\mathcal{W}}^2 \exp(\xi) + \beta \exp(\xi). \end{aligned}$$

Thus we may conclude with the following

THEOREM 4.1. *Under the assumptions of Theorem 3.2, for the matrices D_i defined in (4.1), and $F = \lim_{k \rightarrow \infty} F_k$, where $F_k = \sum_{i=0}^k \frac{1}{i!} D_i$, we have*

$$\begin{aligned} \|D_i\|_{\mathcal{F}} &\leq \sum_{j=0}^{i-1} (\|a\|_{\mathcal{W}} + \|E\|_{\mathcal{F}})^j \gamma_{i-j} \\ \|F\|_{\mathcal{F}} &\leq \left(\frac{1}{2} \|a'\|_{\mathcal{W}}^2 + \|E\|_{\mathcal{F}} \right) \exp(\|a\|_{\mathcal{W}} + \|E\|_{\mathcal{F}}) \end{aligned}$$

where the constants γ_i , $i \geq 1$, are defined in (4.3).

The above theorem states that, also in the case of $A = T(a) + E$, the matrix $F = \exp(A) - T(\exp(a))$ is such that $\sum_{i,j \in \mathbb{Z}^+} |f_{i,j}| < +\infty$. Moreover, if $E = 0$, the bounds given in the above theorem reduce to the bounds given in Theorem 3.3.

With some formal manipulations, it is possible to provide the following bound to $\|D_i\|_{\mathcal{F}}$ expressed in closed form

$$\begin{aligned} \|D_i\|_{\mathcal{F}} &\leq \frac{1}{\|E\|_{\mathcal{F}}} \left(\varphi \frac{(\|a\|_{\mathcal{W}} + \|E\|_{\mathcal{F}})^i - \|a\|_{\mathcal{W}}^i}{\|E\|_{\mathcal{F}}} - \psi i \|a\|_{\mathcal{W}}^{i-1} \right) \\ \varphi &= \|a'\|_{\mathcal{W}}^2 + \|E\|_{\mathcal{F}}^2, \quad \psi = \|a'\|_{\mathcal{W}}^2 \end{aligned}$$

which, taking the limit for $\|E\|_{\mathcal{F}} \rightarrow 0$, coincides with the bound of Theorem 3.3.

5. Algorithms. In this section we provide an algorithm for computing the exponential function of a matrix A such that $A = T(a) + E$ with $a(z), a'(z) \in \mathcal{W}$ and E such that $\|E\|_{\mathcal{F}} < +\infty$.

From the condition $a(z), a'(z) \in \mathcal{W}$ and $E \in \mathcal{F}$ it follows that the coefficients a_i of $a(z)$ decay to zero as $i \rightarrow \pm\infty$, and the entries $e_{i,j}$ of the matrix E decay to zero as $i, j \rightarrow \infty$. Thus we may represent $a(z)$ in an approximate way just by considering a finite number of coefficients, i.e., $a(z) = \sum_{i=-n_-}^{n_+} a_i z^i + r(z)$, where we assume that the remainder $r(z)$ is such that $\|r\|_{\mathcal{W}} \leq \epsilon$, for a given error bound ϵ . Similarly, also the matrix E can be represented in an approximate way by storing only a finite number of nonzero entries. Observe that for the decay of the coefficients a_i , also the matrix $H(a_-)$ can be represented, up to an error ϵ , by means of a semi-infinite matrix which is zero everywhere except in its $n_- \times n_-$ leading principal submatrix, which coincides with the Hankel matrix associated with a_{-1}, \dots, a_{-n_-} .

For computational reasons, it is convenient to represent this semi-infinite matrix as the product UV^T where U and V have infinitely many rows and n_- columns. Moreover, due to the truncation of the series, the matrices U and V have null entries for a sufficiently large row index.

Define $p_i(z) = \frac{1}{i!} a(z)^i$, $i \geq 0$, so that, for a sufficiently large i , $\exp(a)$ can be approximated by $s_i(z)$, defined by means of the recursion

$$\begin{aligned} p_i(z) &= \frac{1}{i} a(z) p_{i-1}(z), \\ s_i(z) &= s_{i-1}(z) + p_i(z), \quad i \geq 1, \end{aligned} \tag{5.1}$$

with $s_0(z) = 1$.

We consider separately, in the following two sections, the Toeplitz case, i.e., $A = T(a)$, and the general case where $A = T(a) + E$ with $E \neq 0$.

5.1. The Toeplitz case. Consider the case where A is Toeplitz, i.e., $A = T(a)$ and $E = 0$. According to the results of the previous section, the matrix $\exp(A)$ is approximated by $T(s_k) + F_k$, for a suitable $k \geq 1$. In order to compute $s_k(z)$ we rely on formula (5.1), while for computing $F_k = \sum_{i=0}^k \frac{1}{i!} E_i$ we rely on recursion (3.2).

We define $\hat{E}_i = \frac{1}{i!} E_i$, so that we may rewrite equation (3.2) in the following form

$$\hat{E}_i = \frac{1}{i} T(a) \hat{E}_{i-1} - \frac{1}{i} H(a_-) H((p_{i-1})_+), \quad i \geq 1,$$

so that $F_k = \sum_{i=0}^k \hat{E}_i$.

In order to reduce the complexity of the computation, we will represent also the matrices \hat{E}_i , F_i and the matrix $H(a_-)$ in the form

$$\hat{E}_i = U_i V_i^T, \quad F_i = W_i Y_i^T, \quad H(a_-) = UV^T, \tag{5.2}$$

where U_i, V_i, W_i, Y_i, U and V are matrices with infinitely many rows and with a finite number of columns. Moreover, due to the finite representation, these matrices have null entries if the row index is sufficiently large. Therefore they can be represented, up to an arbitrarily small error, with a finite number of parameters.

Using the decompositions (5.2) we may write

$$U_i V_i^T = \frac{1}{i} T(a) U_{i-1} V_{i-1}^T - \frac{1}{i} UV^T H((p_{i-1})_+)$$

whence

$$U_i = [T(a)U_{i-1} \quad \mid \quad U], \quad V_i = [\frac{1}{i}V_{i-1} \quad \mid \quad -\frac{1}{i}H((p_{i-1})_+)V]. \quad (5.3)$$

Moreover, from the relation $F_k = F_{k-1} + U_k V_k^T$ we obtain

$$W_k = [W_{k-1} \quad \mid \quad U_k], \quad Y_k = [Y_{k-1} \quad \mid \quad V_k]. \quad (5.4)$$

By using these decompositions, the implementation of equation (3.2), together with the computation of $F_k = \sum_{i=1}^k \widehat{E}_i$ and of the function $s_k(z) = \sum_{i=0}^k p_i(z)$, will proceed as described in the following algorithm, where matrices are formed by a finite number of rows containing the nonzero part of the corresponding infinite matrices.

ALGORITHM 5.1.

INPUT: Integer $k \geq 1$, the coefficient vectors of the functions $a(z)$, $p_{k-1}(z)$, $s_{k-1}(z)$ and the matrices U , V , U_{k-1} , V_{k-1} , W_{k-1} , Y_{k-1} , such that (5.2) holds for $i = k-1$.
 OUTPUT: The coefficient vectors of the functions $p_k(z)$, $s_k(z)$ and the matrices U_k , V_k , W_k , Y_k , such that (5.2) holds for $i = k$.

COMPUTATION:

1. compute $P_1 = H((p_{k-1})_+)V/k$, set $Q_1 = [\frac{1}{k}V_{k-1} \quad \mid \quad -P_1]$
2. compute $P_2 = T(a)U_{k-1}$, set $Q_2 = [P_2 \quad \mid \quad U]$
3. compress the pair Q_1, Q_2 and get a new pair V_k, U_k
4. set $S_1 = [W_{k-1} \quad \mid \quad U_k]$ and $S_2 = [Y_{k-1} \quad \mid \quad V_k]$
5. compress the pair S_1, S_2 and get the new pair W_k, Y_k
6. compute $p_k(z) = \frac{1}{k}a(z)p_{k-1}(z)$ and set $s_k(z) = s_{k-1}(z) + p_k(z)$
7. truncate $s_k(z)$ and $p_k(z)$

In the above description we have used a compression operation in stages 3 and 5 acting on a pair of matrices, together with the operation of truncating a Laurent polynomial at stage 7. We will describe these operations in the next Subsection 5.3.

Observe also that even if the involved matrices have infinitely many rows, only a finite number of them is nonzero. A detailed implementation of the above algorithm should keep track of the number of the nonzero rows of each matrix. We leave this detail to the reader.

5.2. The general case. Consider the case where $A = T(a) + E$, with $E \neq 0$. According to the results of Section 4, the matrix $\exp(A)$ is approximated by $T(s_k) + F_k$, for a suitable $k \geq 1$, where $F_k = \sum_{i=0}^k \frac{1}{i!} D_i$ and the matrices D_i are defined in (4.1).

As in the previous section, define $\widehat{D}_i = \frac{1}{i!} D_i$, so that, in view of (4.2), we have

$$\widehat{D}_i = \frac{1}{i} A \widehat{D}_{i-1} - \frac{1}{i} H(a_-) H((p_{i-1})_+) + \frac{1}{i} E T(p_{i-1}). \quad (5.5)$$

Let us represent the matrices E , $H(a_-)$ and D_i in the form $E = WY^T$, $H(a_-) = UV^T$ and $\widehat{D}_i = U_i V_i^T$, where W , Y , U , V , U_i and V_i have a finite number of columns. We may rewrite equation (5.5) in the form

$$U_i V_i^T = \frac{1}{i} (T(a) + WY^T) U_{i-1} V_{i-1}^T - \frac{1}{i} UV^T H((p_{i-1})_+) + \frac{1}{i} WY^T T(p_{i-1}).$$

Whence we deduce that

$$U_i = [(T(a)U_{i-1} + W(Y^T U_{i-1}) \quad \mid \quad U \quad \mid \quad W],$$

$$V_i = \left[\frac{1}{i} V_{i-1} \quad \mid \quad -\frac{1}{i} H((p_{i-1})_+) V \quad \mid \quad \frac{1}{i} T(p_{i-1})^T Y \right].$$

Moreover, by representing F_k as $F_k = W_k Y_k^T$, from the relation $F_k = F_{k-1} + U_k V_k^T$ we obtain

$$W_k = [W_{k-1} \quad \mid \quad U_k], \quad Y_k = [Y_{k-1} \quad \mid \quad V_k].$$

It is immediate to write an algorithm that implements the above equations. We leave this task to the reader.

5.3. Compression and truncation. Given the matrix E in the form $E = FG^T$ where F and G are matrices of size $m \times k$ and $n \times k$, respectively, we aim to reduce the size k and to approximate E in the form $E = \tilde{F}\tilde{G}^T$ where \tilde{F} and \tilde{G} are matrices of size $m \times \tilde{k}$ and $n \times \tilde{k}$, respectively, with $\tilde{k} \leq k$.

We use the following procedure which, for simplicity, we describe in the case of real matrices. Compute the pivoted (rank-revealing) QR factorizations $F = Q_f R_f P_f$, $G = Q_g R_g P_g$, where P_f and P_g are permutation matrices, Q_f and Q_g are orthogonal and R_f, R_g are upper triangular with columns having non increasing Euclidean norm. Remove the last negligible rows from the matrices R_f and R_g and remove the corresponding columns of Q_f and Q_g . In this way we obtain matrices $\hat{R}_f, \hat{R}_g, \hat{Q}_f, \hat{Q}_g$ such that, up to within a small error, satisfy the equation $F = \hat{Q}_f \hat{R}_f P_f$, $G = \hat{Q}_g \hat{R}_g P_g$. Then, in the factorization $FG^T = \hat{Q}_f (\hat{R}_f P_f P_g^T \hat{R}_g^T) \hat{Q}_g^T$, compute the SVD of the matrix in the middle $\hat{R}_f P_f P_g^T \hat{R}_g^T = U \Sigma V^T$ where the singular values σ_i satisfying the condition $\sigma_i < \epsilon \sigma_1$ are removed together with the corresponding columns of U and V , where ϵ is a given tolerance, say the machine precision. In output, the matrices $\tilde{F} = \hat{Q}_f U \Sigma^{1/2}$, $\tilde{G} = \hat{Q}_g V \Sigma^{1/2}$ are delivered.

This procedure is described with more detail in the next

ALGORITHM 5.2.

INPUT: Matrices F and G of size $m \times k$ and $n \times k$, respectively, a real $\epsilon > 0$;
 OUTPUT: matrices \tilde{F} and \tilde{G} of size $m \times \tilde{k}$ and $n \times \tilde{k}$, respectively, such that $\tilde{k} \leq k$ and $\|FG^T - \tilde{F}\tilde{G}^T\|_2 \leq \gamma \epsilon \|F\|_2 \|G\|_2$ for a suitable $\gamma > 0$.
 COMPUTATION:

1. Compute the pivoted (rank-revealing) QR factorizations $F = Q_f R_f P_f$, $G = Q_g R_g P_g$;
2. select h_f and h_g the smallest integers such that $|(R_f)_{i,i}| < \epsilon |(R_f)_{1,1}|$ for $i > h_f$ and $|(R_g)_{i,i}| < \epsilon |(R_g)_{1,1}|$ for $i > h_g$;
3. denote \hat{R}_f, \hat{R}_g , the submatrices of R_f and R_g formed by the first h_f rows and by the first h_g rows, respectively; denote \hat{Q}_f, \hat{Q}_g the submatrices of Q_f and Q_g formed by the first h_f and h_g columns respectively;
4. compute the SVD of $\hat{R}_f P_f P_g^T \hat{R}_g^T$, i.e., $\hat{R}_f P_f P_g^T \hat{R}_g^T = U \Sigma V^T$;
5. select the integer ℓ such that $\sigma_i < \epsilon \sigma_1$ for $i > \ell$ where σ_i are the singular values, and set \hat{U}, \hat{V} the submatrices formed by the first ℓ columns of U and V , respectively; set $\hat{\Sigma} = \text{diag}(\sigma_1, \dots, \sigma_\ell)$ so that $\|U \Sigma V^T - \hat{U} \hat{\Sigma} \hat{V}^T\|_2 \leq \sigma_{\ell+1}$;
6. output $\tilde{F} = \hat{Q}_f \hat{U} \hat{\Sigma}^{1/2}$, $\tilde{G} = \hat{Q}_g \hat{V} \hat{\Sigma}^{1/2}$.

One can show that the compression obtained this way is such that $\|FG^T - \tilde{F}\tilde{G}^T\|_2 \leq \gamma \epsilon \|F\|_2 \|G\|_2$ where γ is a positive constant depending on the numerical rank.

5.4. Acceleration. The scaling techniques for accelerating convergence of the exponential series described in [14] can be easily implemented in this framework. In

particular, in the case where $A = T(a)$, we determine the least integer q such that $\|a\|_w/2^q < 1$, then we set $\hat{a}(z) = a(z)/2^q$ so that

$$\exp(T(a)) = \exp(T(\hat{a}))^{2^q}$$

and we may compute $\exp(T(a))$ by first computing $\exp(T(\hat{a}))$, which requires a shorter power series expansion, and then computing $\exp(T(\hat{a}))^{2^q}$ by means of q steps of repeated squarings applied to $\exp(T(\hat{a}))$.

The square of a matrix of the kind $T(a) + E$ is computed by means of the equation

$$(T(a) + E)^2 = T(a^2) - H(a_-)H(a_+) + T(a)E + ET(a) + E^2 =: T(a^2) + \hat{E}$$

where $\hat{E} = -H(a_-)H(a_+) + T(a)E + ET(a) + E^2$.

Assuming that E is factored in the form $E = WY^T$, for W and Y being slim matrices, and that $H(a_-)$ is factored as UV^T , being U and V slim, then \hat{E} is factored as $\hat{E} = \hat{U}\hat{V}^T$ where

$$\hat{U} = [-U \mid T(a)W \mid W], \quad \hat{V} = [H(a_+)V \mid Y \mid T(a)^TY + Y(W^TY)]. \quad (5.6)$$

A compression step, performed according to Algorithm 5.2 can be applied to reduce the rank of \hat{U} and \hat{V} . In this case, since some of the involved matrices are Hankel, one can try to get advantage from this property. Observe that, if the numerical rank of the matrix $H(a_+)$ is smaller than that of $H(a_-)$, it's more convenient to express $H(a_+)$ in the form $H(a_+) = UV^T$ where U and V are suitable slim matrices.

The advantage that we obtain by means of the scaling and squaring technique is substantial in many cases.

5.5. Computational analysis. We may perform a complexity analysis of the algorithms designed in the previous sections. Here we give an outline of this analysis and we leave to the reader the completion of the details.

We consider only the case where $A = T(a)$ and divide the problem into the different sub-problems of evaluating the recurrence (3.2) by means of equations (5.3) and (5.4), performing the compression according to Algorithm 5.2, and computing the repeated squaring of a QT matrix.

Concerning (5.3), we have to compute the product $T(a)U_{i-1}$, where $T(a)$ is an infinite Toeplitz matrix having bandwidth $n_- + n_+$, and U_{i-1} has infinitely many rows and a finite number, say r_{i-1} , of columns. Denoting m_{i-1} the number of numerically nonzero rows of U_{i-1} , the problem is reduced to multiplying an $(m_{i-1} + n_-) \times m_{i-1}$ Toeplitz matrix and an $m_{i-1} \times r_{i-1}$ matrix. By using fast algorithms for Toeplitz-vector matrix multiplication we have a cost of $O(r_{i-1}(m_{i-1} + n_-)\log(m_{i-1} + n_-))$ arithmetic operations (ops). Similarly, the computation of the product $H((p_{i-1})_+)V$ is reduced to multiplying an $n_{i-1} \times q_i$ Hankel matrix times a matrix of size $q_i \times s$, where n_{i-1} is the degree of the polynomial $(p_{i-1}(z))_+$, s is the number of columns of V , $q_i = \min(n_{i-1}, n)$, with n the number of numerically nonzero rows of V . Thus, even this computation has a cost of $O(sn_{i-1}\log n_{i-1})$. In fact, the product of a Hankel matrix and a vector, up to permutation, is the same as the product of a Toeplitz matrix and a vector where FFT-based algorithms can be used.

The cost of compression in the steps 3 and 5 of Algorithm 5.1 performed with Algorithm 5.2, which relies on QR and SVD, is proportional to the square of the rank and to the maximum dimension. Finally, the cost of repeated squarings, once the rank-revealing factorization $H(a_-) = UV^T$ has been computed, is dominated by the

compression of the matrices \widehat{U} and \widehat{V} defined in (5.6) which is again proportional to the square of the rank and to the maximum matrix size. On the other hand, the computation of U, V , such that $H(a_-) = UV^T$ involves a rank revealing factorization of the Hankel matrix $H(a_-)$. This factorization is computed by means of a Lanczos-based tridiagonalization with re-orthogonalization which exploits the Hankel structure and the low cost of matrix-vector product. Details on this computation are given in [6] where it is shown that the cost is proportional to the square of the rank and to $n \log n$ where n is the size of the Hankel matrix.

Roughly speaking, the overall cost is proportional to the number of terms in the series, to the square of the maximum rank of the corrections and to the maximum between the band width of $T(a)$ and the size of the correction. In particular, the lower the rank the faster the algorithm.

We may perform a more detailed analysis of the errors generated by the truncation part of the algorithm. This analysis concerns the estimation of the constant γ given in Algorithm 5.2 and it consists of giving technical bounds to the norm of the submatrices involved in the compression by exploiting the properties of rank revealing QR and of SVD. We avoid to provide this analysis and leave it to the reader.

6. Numerical experiments. We have implemented in Matlab the algorithm for computing $\exp(T(a))$, relying on the scaling and squaring acceleration, and performed some numerical experiments with several problems. All the experiments have been run under the Linux system on a I3 processor with Matlab R2016a.

A first bunch of tests concerns the computation of the exponential of symmetric tridiagonal Toeplitz matrices $T = \text{trid}(1, \alpha, 1)$ for different values of α in the range $[-4, 4]$. From the experiments it turns out that the numerical bandwidth of the Toeplitz part, the size and the rank of the correction, as well as the CPU time, are independent of the values of α . The time is roughly 0.01 seconds, the numerical bandwidth is 35 while the correction has size 16×16 and rank 7. The relative norm errors of the approximation are between $3 \cdot 10^{-15}$ and $1.0 \cdot 10^{-14}$.

Another test that we have performed concerns Toeplitz matrices associated with the Laurent polynomial $a(z) = \sum_{i=1}^{n_+} z^i + \sum_{i=0}^{n_-} z^{-i}$ where $T(a)$ is a banded matrix with n_- and n_+ diagonals below and above the main diagonal, respectively. Here the matrix $\exp(T(a))$ has a numerical bandwidth which grows large the larger are n_- and n_+ . In this case, the approximation of $\exp(T(a))$ by means of the finite exponential $\exp(T_n(a))$, where $T_n(a)$ denotes the $n \times n$ leading principal submatrix of $T(a)$, requires that n is as large as at least the numerical bandwidth of $\exp(T(a))$.

We tested the cases obtained with $n_+ = 5$, for different values of n_- where, concerning $T_n(a)$, we have chosen $n = 2m$ where m is the numerical bandwidth of $T(\exp(a))$. We have doubled the value of m in order to remove the boundary effect in the computation of the finite exponential. Table 6.1 reports, the value of n_- , the time t_{QT} needed by our Matlab function to compute $\exp(T(a))$ and the time t_{expm} needed by the Matlab function `expm` for computing the matrix exponential of $T_n(a)$, and the relative error in norm $\|\exp(T(a))_m - \exp(T_n(a))_m\|_\infty / \|\exp(T(a))_m\|_\infty$, together with the numerical bandwidth of $T(\exp(a(z)))$, the size and the rank of the correction F such that $\exp(T(a)) = T(\exp(a)) + F$.

We can see from the table the different growth of the CPU time for the two algorithms and the small value of the rank of the correction, despite the large size of the band. The error is reported only for $n_- \leq 40$ since for larger values of n_- a memory overflow is encountered by the function `expm`.

It is interesting to observe that, roughly, the CPU time grows proportionally to

n_-	t_{QT}	t_{expm}	err	band	rows	cols	rank
10	0.05	0.32	2.3e-14	331	372	245	26
20	0.10	3.52	6.6e-14	831	752	271	23
30	0.10	20.70	2.1e-13	1519	1708	291	18
40	0.12	102.99	2.5e-13	2377	2948	230	11
50	0.15	-	-	3393	3343	214	10
60	0.22	-	-	4563	4490	267	10
70	0.22	-	-	5881	5827	50	9
80	0.29	-	-	7343	7283	49	9
90	0.30	-	-	8947	8867	47	9
100	0.38	-	-	10689	13383	45	8

TABLE 6.1

Exponential of a Toeplitz matrix having 5 diagonals of 1s in the upper triangular part and n_- diagonals of 1s in the lower triangular part. Comparisons of the CPU time needed by the algorithm based on QT matrices and by the Matlab `expm` function. Relative errors in norm of the values computed in the two different ways, band-width of the Toeplitz part, size and rank of the correction F in the matrix exponential $\exp(T(a))$.

the maximum size of the correction and of the bandwidth, times the square of the rank.

In order to figure out how the size and the rank of the correction grow in the partial sum $S_k = \sum_{i=0}^k \frac{1}{i!} A^i$, as k increases, we considered the case where $a(z) = \sum_{j=-20}^{20} a_j z^j$ and the coefficients a_j are randomly distributed in $[0, 1]$. In Figure 6.1 we report the graphs of the number of nonzero rows, the number of nonzero columns and the numerical rank of the correction, as function of k , in the power $T(a)^k$, in the partial sum Y_k and in the matrices generated at any step of the repeated squaring part of the algorithm.

It is interesting to point out that, while the number of nonzero rows and columns in the correction matrices grows with k almost linearly, the value of the numerical rank does not grow much with k . This property makes it cheap to update the correction E_k by means of (5.3) and (5.4), and to perform compression according to Algorithm 5.2.

Another test that we have performed concerns the option pricing problem using the Merton model of [16], [19]. In this case, the exponential of a finite $n \times n$ Toeplitz matrix T_n has to be computed. This matrix can be associated with a symbol $a^{(n)}(z)$ which depends on the value of n . In our experiment, we considered, as test problem, the semi-infinite Toeplitz matrix $T(a^{(n)})$ having as leading principal submatrix the $n \times n$ Toeplitz matrix T_n of the Merton model. We have compared the two $\frac{n}{2} \times \frac{n}{2}$ leading principal submatrices of $\exp(T(a^{(n)}))$ and $\exp(T_n)$ just to verify if the two different matrices have some similarity. We computed $\exp(T_n)$ with the function `expm`, and $\exp(T(a^{(n)}))$ with our algorithm. In Table 6.2 we report the CPU time of this computation and the relative errors in norm concerning the principal submatrices of size $\frac{n}{2}$ of the two matrices, and the rank of the correction F . We can see that the CPU time grows slightly more than linearly with the size n and that the rank of the correction does not grow with n . Also the error, computed only if the size is lower than 4096, takes small values.

7. Final remarks. A classical approach to improve the computation of the matrix exponential is to rely on Padé approximation [14]. This technique consists in approximating the exponential function by means of a ratio of polynomials of low degree. We can apply this technique in our framework provided that an efficient

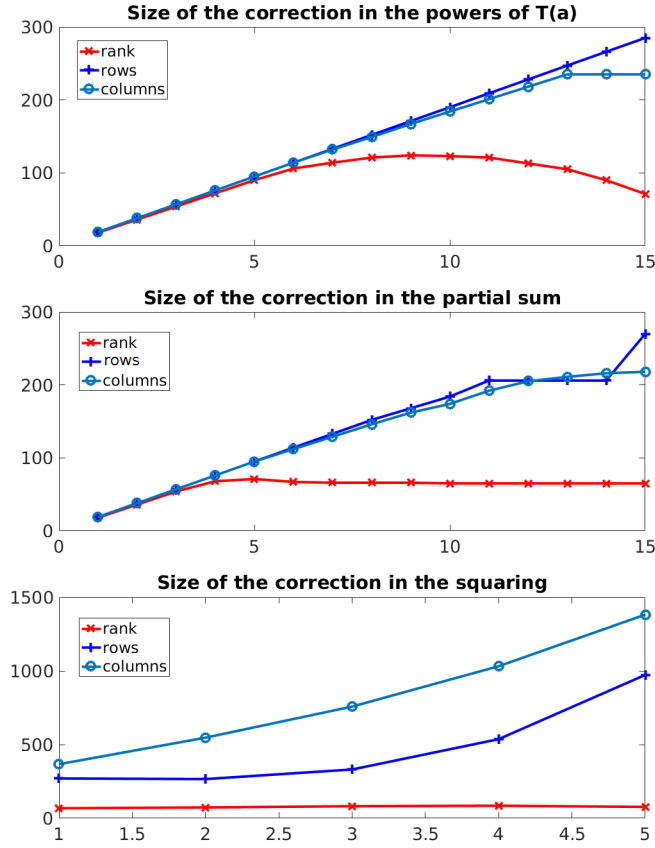


FIGURE 6.1. Growth, as function of k , of the number of rows/columns and of the numerical rank in the correction matrix of $T(a)^k$, S_k , and of the matrices generated by the squaring phase of the algorithm. The function $a(z) = \sum_{i=-20}^{20} a_i z^i$ has random coefficients a_i uniformly distributed between 0 and 1.

n	t_{QT}	t_{expm}	err	rank
512	0.34	0.16	2.7e-12	18
1024	0.72	1.33	2.8e-11	18
2048	2.16	6.43	3.6e-10	18
4096	5.31	-	-	19
8192	17.18	-	-	19

TABLE 6.2
CPU time, error, and rank in the Merton model.

method for computing inverses of QT matrices is designed. This is possible by relying on the Wiener-Hopf factorization of the symbol associated with the Toeplitz part of the matrix. Details on inverting QT matrices are given in [7].

The approach that we have presented in this paper can be applied to deal with general matrix functions and with Toeplitz matrices of finite size n where n is sufficiently large with respect to the decay of the coefficients of $a(z)$. For simplicity consider the case of a Laurent polynomial $a(z) = \sum_{i=-n_-}^{n_+} a_i z^i$ for $n_-, n_+ > 0$, and the associated $n \times n$ matrix $T_n(a) = (t_{i,j})$, $t_{i,j} = a_{j-i}$ if $-n_- \leq j-i \leq n_+$, and

$t_{i,j} = 0$ otherwise. Observe that Theorem 2.1, reformulated for finite matrices and applied to $T_n(a)^2$, leads to the following equation

$$T_n(a)^2 = T_n(a^2) - H_n(a_-)H_n(a_+) - JH_n(a_+)H_n(a_-)J$$

where J is the flip matrix having ones on the anti-diagonal and zeros elsewhere, and $H_n(b)$ denotes the $n \times n$ leading principal submatrix of $H(b)$.

If n is sufficiently large, i.e., $n > n_- + n_+$, then the matrices $H_n(a_-)H_n(a_+)$ and $JH_n(a_+)H_n(a_-)J$ have disjoint supports contained in the upper leftmost corner and in the lower rightmost corner, respectively. Thus, $T_n(a)^2$ can be represented as the sum of the Toeplitz matrix associated with the Laurent polynomial $a^2(z)$ and a correction E which involves a (small) finite number of nonzero entries located in two opposite corners of the support. The same property holds for the powers $T_n(a)^k$ for the values of k for which the size of the numerical support of the correction E does not grow much. Here, for numerical support of a matrix $A = (a_{i,j})$ we mean the set of indices (i, j) for which $|a_{i,j}| < \epsilon \|A\|$ for some norm.

Thus, if the exponential decay of the coefficients of $a(z)$ and of $\exp(a(z))$ is sufficiently large with respect to the size n , then representing the powers $T_n(a)^k$ as well as $\exp(T_n(a))$ in the above form may be computationally effective.

Algorithms for dealing with the finite case can be easily obtained from the algorithms presented in Section 5 by repeating the computation for the correction in the lower rightmost corner involving the matrices $JH_n(a_+)H_n(a_-)J$.

More details in this regard can be found in [6] where it is performed the computational analysis of general functions of finite and infinite Toeplitz matrices expressed either in terms of power series or of Cauchy integrals.

7.1. Conclusions. We have provided a framework to compute matrix functions of a quasi Toeplitz matrix based on the fact that matrices of the kind $A = T(a) + E$ form a matrix algebra if $a(z)$ and $a'(z)$ belong to the Wiener class and $\|E\|_{\mathcal{F}} < +\infty$. A specific analysis has been performed for the exponential function. Numerical experiments confirm the effectiveness of our approach. This framework can be applied to the case of finite matrices and extended to the case of functions expressed by means of a Cauchy integral.

This analysis has set some open issues like analyzing the growth, as a function of k , of the numerical rank of the correction E_k such that $T(a)^k = T(a^k) + E_k$. From the numerical experiments performed so far with some function $a(z)$ this rank seems to be a bounded function of k . Indeed, the growth of the numerical rank of E_k is related to the decay of the coefficients of the function $a(z)$ and it is worth being investigated.

Acknowledgments. Dario Bini wishes to thank Bernd Beckermann and Daniel Kressner for very helpful discussions on topics related to the subject of this paper.

REFERENCES

- [1] B. Beckermann and A. Townsend. On the singular values of matrices with displacement. arXiv:1609.09494, 2016.
- [2] M. Benzi and P. Boito. Decay properties for functions of matrices over C^* -algebras. *Linear Algebra Appl.*, 456:174–198, 2014.
- [3] M. Benzi and V. Simoncini. Decay bounds for functions of hermitian matrices with banded or Kronecker structure. *SIAM J. Matrix Anal. Appl.*, 36(3):1263–1282, 2015.
- [4] D. Bertaccini and F. Di Benedetto. Spectral analysis of nonsymmetric quasi-Toeplitz matrices with applications to preconditioned multistep formulas. *SIAM J. Numer. Anal.*, 45(6):2345–2367 (electronic), 2007.

- [5] D.A. Bini, S. Dendievel, G. Latouche, and B. Meini. Computing the exponential of large block-triangular block-Toeplitz matrices encountered in fluid queues. *Linear Algebra Appl.*, 502:387 – 419, 2016.
- [6] D.A. Bini, S. Massey, and B. Meini. On functions of quasi Toeplitz matrices. Submitted for publication, arXiv:1611.06406, 2016.
- [7] D.A. Bini, S. Massey, and B. Meini. Semi-infinite quasi-Toeplitz matrices with applications to QBD stochastic processes. Submitted for publication, arXiv:1611.06337, 2016.
- [8] A. Böttcher and S. M. Grusky. *Spectral Properties of Band Toeplitz Matrices*. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 2005.
- [9] A. Böttcher and B. Silbermann. *Introduction to large truncated Toeplitz matrices*. Springer Science & Business Media, 2012.
- [10] S. Dendievel and G. Latouche. Approximations for time-dependent distributions in Markovian fluid models *Methodol. Comput. Appl. Probab.*, pages 1–25, 2016. doi:10.1007/s11009-016-9480-0
- [11] G. Fayolle, R. Iasnogorodski, and V. Malyshev. *Random Walks in the Quarter-Plane*. Springer, 1999.
- [12] V. Grimm. Resolvent Krylov subspace approximation to operator functions. *BIT*, 52(3):639–659, 2012.
- [13] P. Henrici. *Applied and Computational Complex Analysis, Vol. 1*. John Wiley & Sons, New York, 1974.
- [14] N. J. Higham. *Functions of Matrices: Theory and Computation*. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 2008.
- [15] M. Hochbruck and A. Ostermann. Exponential integrators. *Acta Numer.*, 19:209–286, 2010.
- [16] D. Kressner and R. Luce. Fast computation of the matrix exponential for a Toeplitz matrix. arXiv 1607.01733v1, 2016.
- [17] S. T. Lee, X. Liu, and H.-W. Sun. Fast exponential time integration scheme for option pricing with jumps. *Numer. Linear Algebra Appl.*, 19(1):87–101, 2012.
- [18] S. T. Lee, H.-K. Pang, and H.-W. Sun. Shift-invert Arnoldi approximation to the Toeplitz matrix exponential. *SIAM J. Sci. Comput.*, 32(2):774–792, 2010.
- [19] R. C. Merton. Option pricing when underlying stock returns are discontinuous. *J. of Financial Economics* 3:125–144, 1976. doi:10.1016/0304-405X(76)90022-2
- [20] A. J. Motyer and P. G. Taylor. Decay rates for Quasi-Birth-and-Death processes with countably many phases and tridiagonal block generators. *Adv. Appl. Prob.*, 38:522–544, 2006.
- [21] H.-K. Pang and H.-W. Sun. Shift-invert Lanczos method for the symmetric positive semidefinite Toeplitz matrix exponential. *Numer. Linear Algebra Appl.*, 18(3):603–614, 2011.
- [22] S. Pozza and V. Simoncini. Decay bounds for non-hermitian matrix functions. arXiv:1605.01595v3, 2016.
- [23] Y. Sakuma and M. Miyazawa. On the effect of finite buffer truncation in a two-node Jackson network. *Stoch. Models*, 12:143–164, 2005.
- [24] S. Serra Capizzano. Generalized locally Toeplitz sequences: spectral analysis and applications to discretized partial differential equations. *Linear Algebra Appl.*, 366:371–402, 2003.
- [25] M. Shao. On the finite section method for computing exponentials of doubly-infinite skew-Hermitian matrices. *Linear Algebra Appl.*, 451:65–96, 2014.
- [26] Y. Takahashi, K. Fujimoto, and N. Makimoto. Geometric decay of the steady-state probabilities in a Quasi-Birth-Death process with a countable number of phases. *Stoch. Models*, 14:368–391, 2001.
- [27] G. Wu, T.-T. Feng, and Y. Wei. An inexact shift-and-invert Arnoldi algorithm for Toeplitz matrix exponential. *Numer. Linear Algebra Appl.*, 22(4):777–792, 2015.